doi: 10.14355/ie.2015.03.001

A Study on Real-Time Business Intelligence and Big Data

Ms.P.R.S.M.Lakshmi[1], Ms.K.SanthiSri[2], Mr.N.Veeranjaneyulu[3]

Department of Information technology, VFSTR University, Vadlamudi

[1]patibandla.lakshmi@gmail.com; [2]srisanthi@gmail.com; [3]veeru2006n@gmail.com

Abstract

Most of the organizations is deploying Business Intelligence (BI) for optimizing every day business operations from past few years. As far as this operational BI approach has been consummate mainly by sinking the latency of data integration and data analysis tasks in traditional enterprise Data Warehousing systems. While sinking these latencies allow earlier to Real-Time (RT) analysis of business operations, it does not support Real-Time decisions to be completed on Real-Time data, i.e., it does not support Real-Time operational BI. Yet, few organizations have recognized their need for RT operational BI. Still anywhere they have, the convolution and costs have often been too lofty. The arrival of big data, though, changes the situation radically. Big data is a shoddily defined and hackneyed term, but, to us, it represents not only new sources of data that aid the analysis and optimization business operations, but also the advances in the erudition and supremacy of analytic techniques. It also reflects the significant advances made by vendors in reducing the costs and improving the performance of analytic software and hardware platforms. Big data increases the number of use cases for RT operational BI and, given the improvements in the price/performance of analytic processing, this makes RT processing viable for a mounting number of organizations.

In this paper, we explore the use cases for RT operational BI in the epoch of big data, examine the technologies that enable RT operational BI, and discuss how the traditional Enterprise Data Warehouse (EDW) environment can be extended to support both RT processing and Big Data.

Keywords

Erudition; Convolution; Shoddily

Introduction

"Big Data" is defined as "Represents the progress of the human cognitive processes, usually includes data sets with sizes beyond the ability of current technology, method and theory to capture, manage, and process the data within a tolerable elapsed time". Recently, the definition of big data as also given by the Gartner: "Big Data are high-volume, high-velocity, and/or high-variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization". According to Wikimedia, "In information technology, big data is a collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools".[1] Many researchers have suggested that commercial DBMSs are not suitable for processing extremely large scale data. Classic architecture's potential bottleneck is the database server while faced with peak workloads. One database server has restriction of scalability and cost, which are two important goals of big data processing.

While processing a query in big data, speed is a significant demand. However, the process may take time because mostly it cannot traverse all the related data in the whole database in a short time. In this case, index will be an optimal choice. At present, indices in big data are only aiming at simple type of data, while big data is becoming more complicated. The combination of appropriate index for big data and up-to-date preprocessing technology will be a desirable solution when we encountered this kind of problems. Application parallelization and divide-and-conquer is natural computational paradigms for approaching big data problems. But getting additional computational resources is not as simple as just upgrading to a bigger and more powerful machine on the fly. The traditional serial algorithm is inefficient for the big data. If there is enough data parallelism in the application, users can take advantage of the cloud's reduced cost model to use hundreds of computers for a short time costs. Rapid advances in computer technology allow business intelligence (BI) systems to provide managers with access to a tremendous amount of data. To function these systems combine complex front-end software with ETL capabilities that extract enormous amounts of data. At the heart of these systems are huge enterprise data warehouses that can populate a possible infinite combination of advanced reports, OLAP cubes and datasets for data mining. The underlying belief is that technically advanced systems

are the most important drivers of effective decision making. Based on this belief BI vendors focus on technologically advanced systems while paying relatively little attention to whether these systems meet the needs of decision makers.

RT Operational Intelligence Business Benefits

All employees are decision-makers at some point during their workday. Front line workers, managers and executives making instant decisions about customers, products, orders and even logistics require the most present information they can obtain to make certain accurate and suitable reactions and results. [2]Suitable information leads to numerous business benefits:

- *Closer decisions* –prevailing at the precise moment with a customer who is about to agitate can save that relationship; shipping a replacement part before the old part fails means less down time.
- *Additional proficient business processes* – identifying concert bottlenecks in a telephone network or web store leads to better customer satisfaction; monitoring and optimizing call center performance reduces customer wait times and enhances call center efficiency.
- *Enhanced revenues and reduced costs* – offering a consumer the correct product at the accurate time results in a sale that might otherwise have been lost; RT analysis can reduce inventory-haulage costs by having just the right amount in stock, and never having an outage.

RT Operational Intelligence Technology Requirements

The objective of RT operational intelligence is to facilitate organizations to build RT decisions on RT data. [5]To understand the technologies required to enable RT processing, it is useful to examine how business intelligence is implemented and used in organizations and then discuss how this is affected by RT operational intelligence. The BI lifecycle diagram in Figure 1 will be used to aid this discussion.

At the center of the diagram are the data sources that provide information about the business operations of the organization. Till date these data sources have consisted mainly of the databases and files created by business transaction processing applications. The use of big data, however, has extended these data sources to include RT event data (from the web environment, sensors, etc.) and data from other internal and external data sources (such as call-center logs, social media sites, external information providers, etc.). These big data sources can add substantial business value to the decision making process, but the challenge of course is extracting this value in a cost-effective and timely manner [4].

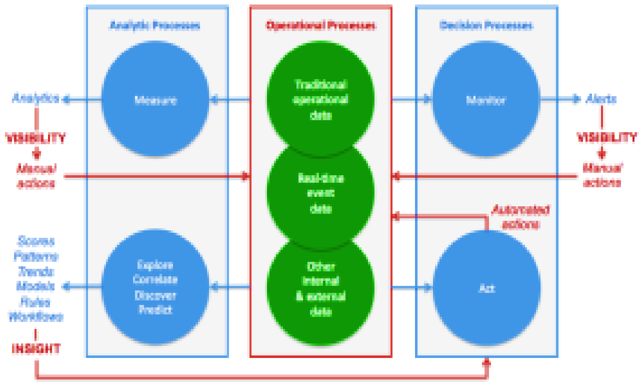

FIGURE 1: THE BUSINESS INTELLIGENCE LIFECYCLE

In a traditional EDW environment, information from operational source systems is captured, transformed and consolidated into an EDW data store ready for analysis. The data in the warehouse represents a historic snapshot of business operations at the point-in-time the data was captured from operational systems. BI applications use this historic information to produce analytics (or measures) about business operations at the time the snapshot was taken. Business users then apply their business knowledge and expertise to the analytics delivered by the EDW system to understand how the business is performing and to take any actions required to optimize business processes or correct any business problems.

The speed of business decisions and actions can be improved by using advanced analytic techniques (such as scoring, data pattern recognition, forecasting, predictive modeling, and statistical and text analysis) to create intelligent analytic objects (such as analytic models, business rules and workflows) that can be embedded in RT decision processes. This approach is illustrated in the bottom half of Figure 1. The decision processes work in conjunction with operational processes to analyze data in real time. This analysis may result in alerts to warn business users about urgent business issues that need immediate attention, or may, in business critical situations such as fraud detection, generate automated actions. The EDW environment can then be used measure the business

impact of the decision processes and, if required, to update the intelligent objects. The need to support faster business decisions and actions, and also to exploit the business benefits of advanced analytic techniques and big data, adds new technology requirements to the traditional EDW environment.

1. Capture and transform real-time event data

2. Capture and transform not only structured data, but also multi-structured data from other internal and external data sources

3. Reduce the latencies of data capture and transformation, analytic processing and user decision-making and action making

4. Support high-performance advanced analytic techniques that enable improved insight in the operation and performance of business processes through the creation of intelligent analytic objects such as scores, models, rules and workflows

5. Embed intelligent analytic objects into decision processes that can analyze data in real-time and generate automated real-time alerts and actions

For example, the more up-to-date the data is in the EDW with respect to operational systems, the better visibility there will be into recent business operations. There is a limit, however, to how up-to-date the data in a data warehouse can be, how fast analytics can be produced and delivered to business users, and how rapidly users can make decisions and take actions. There are inherent latencies in the traditional EDW BI lifecycle, and although these latencies can be reduced, they cannot be eliminated to enable real-time decisions to be made on real-time data. The traditional EDW environment therefore needs to be extended to support RT operational intelligence.

RT Operational Intelligence Support and Big Data

Extended data warehouse architecture for supporting RT operational intelligence and big data is shown in Figure 2. This architecture handles both *in-motion* data and *at-rest* data. In-motion data is real-time data flowing through the operational environment shown at the bottom of the diagram. At-rest data is a snapshot of in-motion data from operational systems, transformed and then loaded into the EDW and investigative computing environments shown at the top of the diagram. The time difference, or data latency, between the at-rest data and real-time is dependent on how often the at-rest data is refreshed or the snapshot taken[4].

The bridge between in-motion data systems and at-rest data systems can be built in two different ways. The traditional method is to use data integration tools and services that use an extract, transaction and load (ETL) approach to bridge the data. A growing trend, however, is to first load the captured in-motion data into a staging area or data refinery, transform it, and then route the transformed data into other systems as required.

Analytic processing can be done by at-rest data systems (such as an EDW or investigative computing platform) or by in-motion data systems in the operational real-time environment or both.

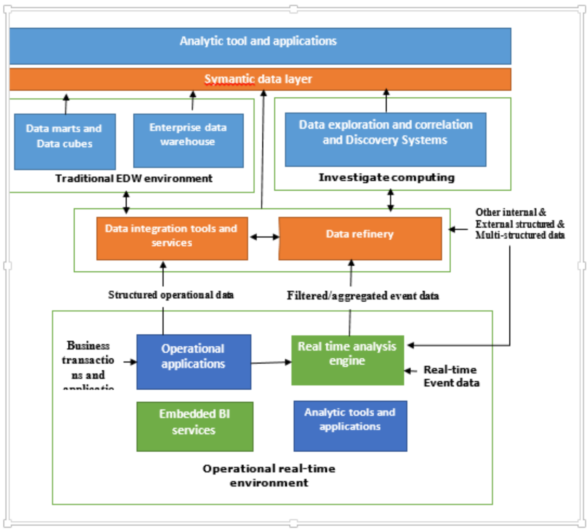

FIGURE2: THE EXTENDED DATA WAREHOUSE ARCHITECTURE

As already discussed, intelligent analytic objects from the at-rest data systems can be embedded in decision processes for real-time data analysis and for automating decisions and actions. These can be implemented as embedded BI services for use within individual operational business processes and applications as shown in the bottom left of Figure 2. One challenge with this approach, however, is that these services are often difficult to embed in existing monolithic business transaction applications. Organizations are therefore more likely to use these BI services when they develop new or reengineered operational applications using a services-oriented architecture and business process management software[5].

Another powerful and relatively new approach to implementing RT operational intelligence solutions is

using a real-time data analysis engine. The technology behind many of these solutions was developed originally for the financial industry for tasks such as analyzing real-time trading. More recently, the technology has been extended to support big data sources and is being used for a variety of applications in a number of industries.

Real-time data analysis engines can process both structured data and multi-structured data. They are particularly well suited to the handling of very large volumes of event streams from sensor devices. Technology terms often associated with a real-time analysis engine are event stream processing (ESP) and complex event processing (CEP). The dividing line between ESP and CEP is blurring as vendors build solutions that support both styles of processing. As a result CEP is becoming the dominant term[7].

ESP is focused on data streams where the events being processed have a definite time sequence. The software provides a windowing capability that enables an application to analyze a data stream for a fixed period of time. So, for example, a financial application can analyze a trading data stream starting in five minutes for a period of ten minutes.

CEP on the other hand is targeted at correlating multiple data streams that may be unrelated to each and may not necessarily be in time sequence. An example would be correlation of retail sales with weather data to determine how the weather is affecting the purchase of certain items. This is a simplification of the ESP and CEP approaches, but the key points to note are that there are various techniques for analyzing real-time data streams and that vendor product now support a number of them.

Real-time data analysis engines can employ the intelligent analytic objects created by data-at-rest systems. They also often provide their own tools for developing such objects. Regardless of their derivation, these objects can be used to do real-time data analysis of real-time data for tasks such as fraud detection, next best customer offer, and so forth. They can also be used to filter high-volume events for use by data refineries and data-at-rest-systems. This is shown in the bottom right of Figure 2.

Companies deploying real-time data analysis engines sometimes use them in conjunction with an investigative computing platform. The real-time engine is used to filter event data for use by the investigative computing platform, which in turn creates intelligent analytic objects for use by the real-time analytic engine for analyzing real-time data. These intelligent analytic objects are built by blending and analyzing the filtered event data and data from at-rest data sources [6].

Real-time data analysis engines support all of the requirements outlined earlier in this paper. They are well suited for building RT operational intelligence applications using both existing data and new big data sources. They also enable a wide range of applications to be developed from churn analysis and fraud detection to equipment monitoring and traffic flow optimization.

The architecture shown in Figure 3 consists of a variety of components that enable companies to extend their traditional EDW environment to support both RT operational intelligence and big data. Most organizations are likely to evolve to such architecture over a period of time. It is unlikely that any single vendor can supply all of these components, and therefore care must be taken when extending the EDW environment to ensure that as components are added they can be integrated easily into the existing IT infrastructure and are sufficiently open to be able to support possible future enhancements to the architecture.

Real-time Operational Intelligence Use Cases

Use cases for RT operational intelligence vary considerably, but, in general, they fit into three broad categories: customer analytics and revenue generation, operational optimization and cost reduction, and asset protection and loss prevention.

Customer Analytics and Revenue Generation

The importance of customer retention and satisfaction cannot be underestimated. Monitoring a highly profitable customers' experiences with your company, its products and personnel in real-time is critical to ensuring their continued loyalty and profitability. Companies are now able to perform 1:1 marketing based on where customers have been, where they are now, and where they are likely to go next. They can detect anomalies in behaviors indicating customers who are likely to churn and quickly offer new or additional products and service plans that fit the customer's lifestyle and usage patterns.

The ability to take instant actions, automate escalations and begin intermediation actions to stop customers

from leaving, improves the generation or retention of the revenue streams from customers [5]. It increases customer satisfaction and gives the organization immediate insight into what works and what doesn't. It also enhances the organization's ability to correct problems before customers even know a problem has occurred.

Operational Optimization and Cost Reduction

In today's economy, corporations everywhere must squeeze inefficiency out of all their operational processes as quickly as they can to avoid higher costs and reduce wastefulness. The ability to detect these inefficiencies in real time is another good use case for RT operational intelligence. An example here is supply chain management. Often a company runs the risk of stock-outs due to a misalignment between customer demand for a product and the company's inventory resulting in significant revenue loss as well as customer dissatisfaction. Being able to detect these problems early reduces missed shipments, overages and underages and reduces costly emergency replenishment due to process inefficiency. As another example, certain types of complex equipment and devices such as cell towers, networks, aircraft, oil wells and automobiles often provide "signals" of pending failures. Unfortunately many organizations are not able to recognize these signals, which results in avoidable outages and unhappy customers. The good news is that companies now have the technology to determine potential product outages or failures affecting their customers in real time before they actually happen.

Asset Protection and Loss Prevention

The most obvious application for RT operational intelligence in this category is being able to detect fraudulent transactions within seconds of their creation. The first step is to develop models of fraudulent behavior from a set of known fraudulent transactions. These models are used in the RT operational intelligence applications to compare to new transactions entering the systems. Those that match the fraudulent characteristics in the models are then redirected to another set of automated or manual processes for further investigation.

Financial companies are also using sophisticated algorithms and analyses for early detection of "out of process" outliers and suspicious activity. They can detect and predict transactions that will be in jeopardy of default (car and home loans, transactions missing documentation, for example) resulting in fewer late settlements, fewer fines and improved regulatory compliance. They can also use early detection analyses to determine potential cyber security breaches as well. There are many other use cases for RT operational intelligence involving situational awareness, smart call center management and service level agreement monitoring and management. In the second of these papers, we will expand on these use cases by examining detailed customer case studies.

Conclusion

Many corporations have believed that RT operational intelligence was further than their technological capabilities as well as their budget. The good news is that this is no longer true. Today, the complexity of the environment has been greatly simplified and its total costs reduced. We can give a fair amount of credit to the big data push in many organizations for the dramatic increase in the demand for RT operational intelligence.

RT operational intelligence has also demonstrated significant benefits to many of today's modern corporations. The ability to make faster, more accurate decisions based on real-time data has been proven to preserve revenue streams, decrease costs and minimize significant risks. But the creation of this environment should be an evolutionary process.

Most organizations move toward RT operational intelligence through a progression of BI maturity – from simple measurements to sophisticated insight to real-time action. Many companies start by speeding up the data integration processes in their traditional enterprise data warehouse environments. The ability to reduce the latency of structured data certainly moves the organization closer to operational intelligence, but it still a far cry from RT operational intelligence.

The enterprise must squeeze the extended data warehouse architecture to fully achieve the real-time analytical capabilities described in this paper. This architecture includes both the exploratory computing environment and, more importantly, the analytic capabilities found within the operational environment itself – embedded BI services and a real-time analysis engine.

The use cases for RT operational intelligence fall within the broad categories of customer analytics and revenue generation, operational optimization and cost

reduction, and asset protection and loss prevention. Every industry and government entity can easily generate their business cases within these to justify the implementation of their own RT operational intelligence capabilities.

REFERENCES

"Big data: science in the petabyte era," *Nature 455 (7209):1*, 2008.

Douglas and Laney, "The importance of 'big data': A definition,"2008.

Gardner, Dana. "GigaSpaces Survey Shows Need for Tools for Fast Big Data, Strong Interest in Big Data in Cloud." *ZDNet BriefingsDirect* (December 10, 2012). zdnet.com/ gigaspaces-survey-showsneed-for-tools-for-fast-big-data-strong-interest-in-big-data-incloud-7000008581/

Groenfeldt, Tom. "Big Data—Big Money Says It Is a Paradigm Buster."*Forbes* (January 6, 2012). forbes.com/ sites/tomgroenfeldt/2012/01/06/big-data-big-money-says-it-is-a-paradigm-buster/

Peer Research: Big Data Analytics: Intel's IT Manager Survey on How Organizations Are Using Big Data. Intel® IT Center (August 2012).

Peer Research: Mission-Critical Workload Migration. Intel (January 2013). intel.com/content/www/us/en/risc-migration/mission-critical-workloadmigration-peer-research-report.html

X. Zhou, J. Lu, C. Li, and X. Du, "Big data challenge in the management perspective," *Communications of the CCF*,vol. 8, pp. 16–20, 2012.